AL-TP-1991-0057

AD-A245 099

ARMSTRONG LABORATORY

GENERAL COGNITIVE ABILITY PREDICTS JOB PERFORMANCE

Malcolm James Ree
James A. Earles

HUMAN RESOURCES DIRECTORATE
MANPOWER AND PERSONNEL RESEARCH DIVISION
Brooks Air Force Base, TX 78235-5000

Mark S. Teachout

HUMAN RESOURCES DIRECTORATE
TRAINING SYSTEMS RESEARCH DIVISION
Brooks Air Force Base, TX 78235-5000

January 1992

Interim Technical Paper for Period November 1990 – November 1991

Approved for public release; distribution is unlimited.

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5000

NOTICES

This technical paper is published as received and has not been edited by the technical editing staff of the Armstrong Laboratory.

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

Malcolm James Ree

MALCOLM JAMES REE
Scientific Advisor

William E. Alley

WILLIAM E. ALLEY, Ph.D
Technical Director
Manpower and Personnel Research Division

Roger W. Alford

ROGER W. ALFORD, Lieutenant Colonel, USAF
Chief, Manpower and Personnel Research Division

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	January 1992	Interim – November 1990 – November 1991

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
eneral Cognitive Ability Predicts Job Performance	PE – 62205F PR – 7719 TA – 18 WU – 67
6. AUTHOR(S) Malcolm James Ree James A. Earles Mark S. Teachout	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Armstrong Laboratory Human Resources Directorate Manpower and Personnel Research Division Brooks Air Force Base, TX 78235 5000	8. PERFORMING ORGANIZATION REPORT NUMBER AL-TP-1991-0057
9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)	10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT	12b. DISTRIBUTION CODE
Approved for public release; distribution is unlimited.	

13. ABSTRACT (Maximum 200 words)

The roles of general ability (**g**) and specific abilities (**s**) were investigated as predictors of several job performance criteria for Air Force enlistees in 8 jobs. General cognitive ability and specific abilities (the interaction of general ability and experience) were defined by scores on the first and subsequent principal components of the enlistment selection and classification test, the Armed Services Vocational Aptitude Battery. Multiple regression conducted in correlation matrices corrected for range restriction revealed that **g** was the best predictor of all criterion measures and that **s** added a statistically significant but practically small amount to predictive efficiency. For classification of large numbers of applications into large numbers of jobs the incremental prediction due to **s** could be useful.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Armed Services Vocational Aptitude Battery (ASVAB) validity cognitive ability factor analysis	30
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

TABLE OF CONTENTS

LIST OF TABLES

PREFACE

This effort was conducted under work unit 77191867, Research on Air Force Selection and Classification. The authors thank SSgt David Lebrun for his computer analyses. Gratitude is extended to Mr. Bill Glasscock for his consultation. For critical reading of early drafts, the authors thank Thomas Watson and Lonnie D. Valentine, Jr., both of AL/HRMI and Terry Dickinson, Old Dominion University. William Alley, AL/HRM, Arthur Jensen, University of California at Berkeley, and Howard Wainer, Educational Testing Service (ETS), are owed debts of gratitude for insightful discussions on the topic.

"When a thing ceases to be a subject of controversy,
it ceases to be a subject of interest."

William Hazlitt

SUMMARY

Many multiple aptitude test batteries, including the Armed Services Vocational Aptitude Battery (ASVAB), used for assigning or classifying individuals to jobs or for occupational counseling have subtests covering a broad range of content such as science, mathematics, reading, vocabulary, clerical, mechanical, or technical knowledge. This content reflects a belief that job performance is best predicted by subtests whose content appears to be closely related to the tasks of the job. It has been demonstrated that the subtests of a multiple aptitude test battery all measure, in large part, general learning ability in addition to the specific abilities implied by the differing contents of the subtests.

This study investigated the utility of general learning ability and specific abilities for predicting job performance criteria in eight Air Force jobs. Subjects were 1,545 Air Force enlistees. It was found that general ability was the best predictor and that specific abilities improved the predictive accuracy by a small amount. However, small increments can be useful in classification when large numbers of applicants are available to be assigned to large numbers of jobs.

GENERAL COGNITIVE ABILITY PREDICTS JOB PERFORMANCE

I. INTRODUCTION

The concept of general cognitive ability or psychometric **g** first proposed by Galton in 1883 appeared in analyses early in this century. Spearman (1904) proposed a two factor theory of abilities including general cognitive ability, **g**, and specific abilities, **s**. The relative importance of **g** and **s** in the prediction of criteria has been and remains the center of controversy.

Early intelligence test developers such as Binet and Simon were proponents of **g** but eventually the influence of multiple ability theorists such as Thurstone (1938) was pervasive. This led to the development of multiple aptitude test batteries. The Differential Aptitude Tests (DAT), the General Aptitude Test Battery (GATB), and the Armed Services Vocational Aptitude Battery (ASVAB) were designed to measure specific abilities and to make specific predictions about employment or education. Sets of test scores would be differentially selected or differentially weighted for each situation, fulfilling a proposal by Hull (1928). It was proposed that specific abilities could compensate for a lack of general ability. The different composites of subtests used by the military for job placement or the interpretation of score profiles in counseling are current examples of the application of multiple ability theory. The use of differential weighting and different composites led to multiple aptitude theory being termed a theory of "differential validity" (Brogden, 1951).

Jensen (1980) has identified **s** with specific experience rather than with specific ability. In this same vein, Cattell (1971; 1987) posited his "investment theory" which proposes that initially there is a general ability, (called fluid **g** or g_f) which is invested in specific experiences and crystallizes to specific skills (called crystallized **g** or g_c). This means that **s** is **g** modified by experience. It implies that for an individual, the best estimate of **g** can be made from testing content in which the individual has invested their ability (g_c) or from tests which require little or no prior special experience (g_f) from training, interest, motivation, or exposure. An example of the former is that unsatisfactory estimates of **g** would be obtained by administering a French test to a sample half of which has studied French and half of which has not. The estimates of **g** for the half which did not study French would be unsatisfactory; the estimates for the other half would be more satisfactory. To rectify this problem, Raven (1938), a student of Spearman's, developed his Progressive Matrices test which measured **g** through a series of "abstract diagrammatic problems" (Vernon, 1960, p. 19) which did not require special investment of **g** but rather that **g** be used to solve nonverbal problems.

The primacy of **g** as a predictor has again become the subject of many studies. The December 1986 issue of Journal of Vocational Behavior (Gottfredson, 1986) documented the renewed interest as did the evidence emerging from validity generalization studies (Hunter, 1983, 1984a, 1984b, 1984c; Hunter, Crossen, & Friedman, 1985).

The ASVAB is an excellent source of data for investigating the value of g as a predictor, with over one million administrations and over 200,000 selections to job training each year.

Jones (1988) correlated the average validity of the ASVAB subtests for predicting training performance with the g saturation of the subtests. For each subtest, the corrected for range restricted training validities were averaged over 37 diverse Air Force technical training courses. These averages were subject weighted over a total of 24,482 technical training students. For each subtest, the g saturation was measured by its loading on the unrotated first principal component (see Jensen, 1987; Ree & Earles, 1991a). She found a rank-order correlation of .75, demonstrating a strong positive relationship between g and predictive efficiency. This was found across all jobs and comparable values were found within the four Air Force job families of Mechanical, Administrative, General Technical, and Electronics. Following Jensen (1980), the Jones rank-order correlation was calculated for the all job condition as .98 after correcting the g loadings for subtest unreliability.

Ree and Earles (1990) investigated the predictive utility of both the general and specific components of the ASVAB by regressing Air Force technical school grades on the unrotated principal component scores of the ASVAB. Psychometric g was represented by the first principal component and special or invested abilities by the remaining principal components. Across 89 jobs (individual sample sizes ranged from 274 to 3,939), the average correlation of g and the training criterion was .76 corrected for range restriction. When the specific (g x experience) components were added to the regressions, the R increased an average of .02.

Using a linear models approach, Ree and Earles (1991b) evaluated the nature of the relationships of g and specific or invested abilities to 82 Air Force job training criteria. They found statistically significant, but practically trivial, contributions (an average gain of .02) of specific or invested abilities to the regressions.

These three studies examined the predictive utility of general ability (and two the contribution of specific abilities), but none used job performance measures as criteria. Jones (1988) observed that measures of job performance were the preferred criteria but hardly ever available, frequently due to costs.

In as far as individuals sort themselves into jobs on the basis of their ability to perform, job incumbency becomes a form of job performance. Psychometric g as measured by the Army General Classification Test (AGCT)(Stewart, 1947) was found to be related to pre-service occupation of soldiers during World War II. Among the jobs with highest average estimated intelligence were accounting, engineering, and medicine. Jobs with middling average estimated intelligence were policeman, electrician, and meat cutter. Jobs with the lowest average estimated intelligence included laborer, farm worker, and lumberjack. The distribution of within job intelligence scores did not overlap for the very highest and very lowest jobs. This study did not consider special or invested abilities.

Hunter (1986) reviewed hundreds of studies which showed that **g** predicted job performance criteria including training success, supervisory ratings, and content valid hands-on work samples for both civilian and military jobs. However, direct tests of the incremental contribution of specific abilities for the prediction of job performance criteria were not made.

An advantage of the current study was the availability of several measures of job performance and measures of both **g** and **s**. The Air Force developed a job performance measurement system that included a work sample, an interview of job procedures, and a supervisory rating of job proficiency. The current study sought to determine if measures of **g** and **s** were differentially (Brogden, 1951) useful predictors of job performance criteria.

II. METHOD

Subjects

The subjects were 1,545 nonprior service Air Force enlistees entering from 1984 through 1988 who had tested with ASVAB parallel forms 11, 12, or 13, had completed both basic military training and technical training and were for the most part, working in their first term of enlistment. They were mostly White (78.1%), male (83.2%), 17 to 23 years old, high school or better graduates (99.1%) with an average job tenure of about 28 months.

Predictors

The Armed Services Vocational Aptitude Battery is a multiple aptitude test battery (DOD, 1984) composed of ten subtests as shown in Table 1. Except for the Numerical Operations and Coding Speed subtests which are speeded, all are power tests. It is used for enlistment qualification and initial job assignment. The battery was normed on a sample of 18-to 23-year-old youths weighted to be nationally representative (Maier & Sims, 1986; Ree & Wegner, 1990). The ASVAB has been used in this subtest configuration since 1980. Its reliability has been studied (Palmer, Hartke, Ree, Welsh, & Valentine, 1988), and it has been validated for many military occupations (Earles & Ree, in press; Welsh, Kucinkas, & Curran, 1990; Welsh, Trent, Nakasone, Fairbank, Kucinkas, & Sawin, 1990; Wilbourn, Valentine, & Ree, 1984).

The Air Force aggregates the subtests into four composites in a reified belief in differential validity. These composites are Mechanical (M = MC + GS + 2AS), Administrative (A = NO + CS + WK + PC), General-Technical (G = WK + PC + AR), and Electronics (E = AR + MK + EI + GS).

Table 1. Subtests of the ASVAB

Subtests	Number of Items	Time in Minutes	Reliability
General Science (GS)	25	11	.80
Arithmetic Reasoning (AR)	30	36	.87
Word Knowledge (WK)	35	11	.88
Paragraph Comprehension (PC)	15	13	.67
Numerical Operations (NO)	50	3	.72
Coding Speed (CS)	84	7	.77
Auto and Shop Information (AS)	25	11	.82
Mathematics Knowledge (MK)	25	24	.84
Mechanical Comprehension (MC)	25	19	.77
Electronics Information (EI)	20	9	.71

Note. Test-retest reliability estimates taken from Palmer et al. (1988).

There are three generally accepted ways of estimating the g component of a set of variables (Jensen, 1980). Ree and Earles (1991) have shown that for the ASVAB, estimates of g from these three methods, principal components, principal factors, and hierarchical factor analysis, all correlated greater than .996. Because of high correlations among the various g estimates and the mathematical simplicity of the principal components, they were chosen to represent the general, g, and specific (or invested), s, measures of the ASVAB. The first unrotated principal component serves as a measure of g (Jensen, 1980). Specific abilities are often represented by group factors from common factors analyses with g ineluctably distributed through them from rotation (Jensen, 1980). The g can be removed from the lower order factors through the Schmidt-Leiman (1955) procedure. However, common factors procedures do not account for all the variance in the variables and put the specific variances at a relative disadvantage compared to principal components procedures which do account for all the variance and provide maximum advantage for the specific abilities.

To determine the maximal predictive efficiency (Brogden, 1946) of the specific abilities, the best choice is the procedure which most fully represents the non-g portions. Therefore, the nine remaining unrotated principal components were used as the measures of specific or invested abilities (s_1 to s_9). These are mathematically defined measures of specific abilities and do not necessarily represent identifiable or namable concepts. Jones (1988) investigated the second principal component and found it to be gender related. When a variable for gender was included in the principal component analysis, it loaded highest on the second component by a considerable amount. If the invesment theory holds this principal component which positively weights the two subtests which female means exceed male means and negatively weight subtests where male means exceed female means is an expression of differential investment.

The principal components have the additional benefit of being orthoginal (Hotelling, 1933a, 1933b) which, according to Kendall Stuart and Ord (1983), avoids the problems of colinearity and enhances their usefulness in regression.

Tables 2 and 3 present the principal component score weights and principal component loadings.

Table 2. Principal Component Weights for the ASVAB Subtests

	Principal Components				
	1	2	3	4	5
GS	.13808	-.11244	-.21982	-.29416	.19523
AR	.13715	.03854	-.39912	.54694	-.02066
WK	.13736	.06649	-.21381	-.64261	-.08976
PC	.12778	.16656	-.31273	-.71570	-.02359
NO	.11291	.38342	.42663	.23843	-1.36760
CS	.09956	.44464	.75816	.03679	1.11560
AS	.10878	-.43374	.60474	-.00918	-.34001
MK	.12965	.12086	-.61486	.64452	.20353
MC	.12448	-.30623	.21087	.39938	.36281
EI	.12857	-.29635	.14351	-.13640	-.00001
	6	7	8	9	10
GS	-.88893	-1.05107	.56764	.46367	-1.25618
AR	.26159	.58641	.25640	-1.51740	-1.06178
WK	-.20343	-.35471	.19392	-1.22910	1.53259
PC	1.10958	.48914	-.18581	.83254	-.55741
NO	-.11449	-.39672	-.29306	.20266	-.11527
CS	-.14894	.21734	.13184	-.06193	-.04099
AS	.22086	.62982	1.28388	.27471	.26269
MK	-.26607	.28551	.29615	1.16925	1.09690
MC	.89768	-1.19071	-.72807	-.02996	.28081
EI	-.78167	.90823	-1.43032	.09391	-.06884

Note. Weights computed in the ASVAB normative sample. See Ree and Earles (1990).

Table 3. Unrotated Principal Components Loadings for ASVAB Subtests

	Principal Component									
	1	2	3	4	5	6	7	8	9	10
GS	.88	-.14	-.11	-.14	.05	-.24	-.22	.11	.07	-.18
AR	.87	.04	-.20	.27	.00	.07	.12	.05	-.24	-.15
WK	.87	.08	-.11	-.32	-.02	-.05	-.07	.03	-.19	.22
PC	.81	.21	-.16	-.36	.00	.29	.10	-.03	.13	-.08
NO	.72	.49	.22	.12	-.39	-.03	-.08	-.06	.03	-.01
CS	.63	.57	.39	.01	.32	-.04	.04	.02	.00	.00
AS	.69	-.55	.31	.00	-.09	.05	.13	.26	.04	.03
MK	.82	.15	-.32	.32	.05	-.07	.06	.06	.18	.16
MC	.79	-.39	.11	.20	.10	.24	-.25	-.14	.00	.04
EI	.82	-.38	.07	-.06	.00	-.21	.19	-.29	.01	-.01

Note. Loadings computed in the ASVAB normative sample. See Ree and Earles (1990).

Jobs

Eight jobs were selected to be representative of all Air Force jobs. Each job had a minimum requirement on one of the four composites. Jet Engine Mechanic and Aerospace Ground Equipment Specialist were selected by the **M** composite; Information Systems Operator and Personnel Specialist by the **A** composite; Air Traffic Controller and Aircrew Life Support specialist by the **G** composite; and Precision Measurement Equipment Specialist and Avionics Communications Specialist by the **E** composite.

Criteria

The criteria were developed as part of the Joint-Services Job Performance Measurement Project (Wigdor & Green, 1987). The measures used in the present study were hands-on-work samples (**HOPT**), technical interviews (**INT**) in which the subjects explained how to perform technical tasks, and the combination of **HOPT** and **INT** called a Walk Through Performance Test (**WTPT**) (Hegde & Lipscomb, 1987). A secondary or surrogate measure was task ratings by supervisors (**SUPR**). Because the WTPT was expensive to develop and administer the surrogate was included in an effort to obtain measures of job proficiency at a lower cost.

Work sample criterion development. A hands-on work sample test (HOPT) was constructed for each job to assess proficiency on representative job tasks. The task domains for each job were identified and defined from the Air Force Occupational Survey data base (Christal, 1974). A domain sampling plan was developed (Lipscomb, 1984), and tasks were sampled with stratified random sampling procedures (Lipscomb, 1984; Lipscomb & Dickinson, 1988).

For each task, work sample developers used technical descriptions of work procedures (Air Force technical orders and manuals) as well as input from subject matter experts (SMEs) to define and describe the procedural steps required for successful task completion. A hands-on work sample test was constructed for each task, reviewed by SMEs, and field tested at several Air Force bases. A "yes/no" format was used to rate the performance on each procedural step within the task. The proportion of steps performed correctly was calculated for each task and this value was the score. Each job had multiple tasks.

Work sample administrator training. The work sample tests were administered to the subjects and rated by active duty or retired noncommissioned officers with extensive job experience. The raters received one to two weeks of scorer accuracy training and observation (Hedge, Lipscomb, & Teachout, 1988). Videotapes of work sample test performance with known target ratings were used as training devices. After viewing and rating the videotapes, the administrators discussed the key work behaviors to perform or avoid for successful task completion. Hedge, Dickinson, and Bierstedt (1988) reported that this training produced accurate and reliable work sample test rating. The raters demonstrated high average agreement ($r = .81$) and high average correlational accuracy ($r = .85$) between their ratings and videotape target ratings.

In addition, a "shadow scoring" technique was used during a portion of data collection with 58 subjects which required two test administrators to observe and rate task performance. The technique was effective in maintaining agreement in the scoring of the work sample tests. The average scorer-shadow scorer agreement was 95% across the 58 subjects.

Supervisory ratings. Graphic rating scales were developed to measure technical proficiency on the same tasks measured by the Walk Through Performance Test. Each task was described by its statement from the Air Force Occupational Survey. Task performance was rated on a 5-point adjectivally anchored scale.

Supervisory ratings training. In a group rater orientation session, the project was described, participation conditions explained, and rating measures presented. This orientation was followed by one hour of frame-of-reference and rater error training (McIntyre, Smith, & Hassett, 1984). Two rating exercises facilitated use of rating forms by identifying varying levels of performance and their associated rating-scale anchors. Participants practiced rating the performance of incumbents described in the two exercises. Following these ratings, they received target-score accuracy feedback. In addition, a third exercise highlighted rating errors, and how to improve rating accuracy.

Procedures

Data collection. Criterion data were collected as part of a project to validate selection and classification tests (Hedge & Teachout, 1986). Immediately following rater training, rating booklets were distributed, and the supervisors completed the rating fonns. Subsequent to the group session, job incumbent subjects were individually administered the WTPTs. Time limits were specified for each WTPT, ranging from four to seven hours.

Analyses. Each criterion was regressed against the set of principal components for each job in a forward stepwise manner with no order of inclusion specified. This was accomplished for the correlations artifactually depressed by prior selection. To make better estimates of the correlations in the unrestricted population, the regressions were also computed in matrices after multivariate correction for range restriction. The Type I error rate was set at $p < .01$.

The F-test statistic for regression (stepwise and non-stepwise) uses the error sum of squares in computation (Ward & Jennings, 1973). The equation below compares two linear models such as is done to determine if another variable can be added in stepwise regressions or when a regression is tested to determine if it is significantly different from zero. There is an alegbraicly equivalent variant of this equation using R^2s, but the ratio remains the same and the F value remains the same.

$$F=((ESS_1-ESS_2)/(df_1-df_2))/(ESS_2/df_2)$$

Where ESS_1 is the error sum of squares for the restricted model and ESS_2 is the error sums of squares for the full model and these are divided by their respective degrees of freedom.

One of the assumptions of the correction for range restriction is that the variance error of estimate (alternate name for the error sum of squares divided by degrees of freedom) are equal in the uncorrected and the corrected regression. The error sum of squares does not change from application of the correction. The computation of the F in the restricted sample and in the corrected sample is therefore algerbaic equivalent. This allows for the computation of the F test for significance of a difference of regressions between two models in the corrected matrices.

III. RESULTS AND DISCUSSION

These analyses disclosed that the principal components were useful in predicting the criteria, as found for training criteria (Ree & Earles, 1990; Ree & Earles, 1991b). The specific or invested abilities, as represented by the second through tenth principal components, added to the accuracy of prediction, but by a small amount. The efficiency of the predictors (uncorrected correlations; r and R) in this study were smaller than in a previous study (Ree & Earles, 1990). The sample sizes in this study were much smaller so that some portion of the increases due to specific or invested ability are likely to be the results of overfitting and likely to diminish on cross-validation. These regression results are reported in Table 4. The correlations within parentheses are estimates of cross-validation coefficients by use of Stein's expectancy operator (see Kennedy, 1988).

Table 4. Correlations and Regressions of Measures of g and s With the Criteria

AFSC 122X0 (n = 162) Aircrew Life Support Specialist

Criteria					Order of Entry of Principal Components in Equation	
	r_g	r_g^c	R_{g+s}	R_{g+s}^c	Uncorrected	Corrected
HOPT		No significant correlations.				
INT			-.28*	-.26	2	2
			(-.24	-.22)		
WTPT		No significant correlations.				
SUPR		.24				1
		(.19)				

AFSC 272X0 (n = 164) Air Traffic Control Operator

Criteria					Order of Entry of Principal Components in Equation	
	r_g	r_g^c	R_{g+s}	R_{g+s}^c	Uncorrected	Corrected
HOPT		No significant correlations.				
INT		.25		.33		1, 5
		(.21)		(.28)		
WTPT		.26				1
		(.22)				
SUPR		.23				1
		(.18)				

Table 4. (Cont'd)

AFSC 324X0 (n = 126) Precision Measuring Equipment Specialist

Criteria	r_g	r_g^c	R_{g+s}	R_{g+s}^c	Order of Entry of Principal Components in Equation: Uncorrected	Corrected
HOPT	.34	.69	.45	.76	1,4	1,4
	(.30	.68	.41	.74)		
INT	.32	.75			1	1
	(.28	.74)				
WTPT	.36	.71	.45	.77	1,4	1,4
	(.32	.70	.41	.75)		
SUPR		No significant correlations.				

AFSC 328X0 (n = 74) Avionics Communications Specialist

Criteria	r_g	r_g^c	R_{g+s}	R_{g+s}^c	Order of Entry of Principal Components in Equation: Uncorrected	Corrected
HOPT	.34	.72	.72	.75	1	1,8
	(.27	.71		.72)		
INT	.26	.61	.41	.73	8,1	1,8,3
	(.16	.58	.32	.69)		
WTPT	.34	.71	.55	.81	1,3,8	1,8,3
	(.27	.69	.48	.78)		
SUPR		.36		.55		1,6,8
		(.30)		(.48)		

Table 4. (Cont'd)

AFSC 423X5 (n = 211) Aerospace Ground Equipment Specialist

Criteria	r_g	r_g^c	R_{g+s}	R_{g+s}^c	Order of Entry of Principal Components in Equation: Uncorrected	Corrected
HOPT	.29	.42	.41	.57	5,1,2	1,2,5
	(.26	.40	.37	.54)		
INT	.19	.31	.28	.45	2,1	1,2,5
	(.14	.28	.23	.41)		
WTPT	.26	.38	.39	.53	5,1,2	1,2,5
	(.23	.36	.35	.50)		
SUPR		No significant correlations.				

AFSC 426X2 (n = 178) Jet Engine Mechanic Specialist

Criteria	r_g	r_g^c	R_{g+s}	R_{g+s}^c	Order of Entry of Principal Components in Equation: Uncorrected	Corrected
HOPT		.25		.31		1,5
		(.21)		(.26)		
INT	.25	.43		.47	1	1,2
	(.21	.41)		(.44)		
WTPT	.19	.35			1	1
	(.14	.32)				
SUPR				.20		4
				(.15)		

Table 4. (Cont'd)

AFSC 492X1 (n = 111) Informations Systems Radio Operator

Criteria	r_g	r_g^c	R_{g+s}	R_{g+s}^c	Order of Entry of Principal Components in Equation: Uncorrected	Corrected
HOPT		.28	-.30	.40	3	3,1
		(.22	-.21	.34)		
INT	.32	.32		.40	1	1,3
	(.27	.27)		(.34)		
WTPT	.27	.34	.37	.44	1,3	1,3
	(.21	.30	.31	.39)		
SUPR	.28	.50			1	1
	(.22	.47)				

AFSC 732X0 (n = 172) Personnel Administration Specialist

Criteria	r_g	r_g^c	R_{g+s}	R_{g+s}^c	Order of Entry of Principal Components in Equation: Uncorrected	Corrected
HOPT	.22	.49			1	1
	(.17	.47)				
INT		.46		.50		1,2
		(.44)		(.47)		
WTPT	.21	.53		.56	1	1,2
	(.16	.51)		(.54)		
SUPR		No significant correlations.				

Note. Correlations with superscripts c indicate correction for range restriction. Correlations without superscripts are as observed. Correlations within parentheses are cross-validation estimates using Stein's expectancy operator. r_g is the correlation of g and the criterion, r_g^c is the correlation between g and the criterion corrected for range restriction, R_{g+s} is the multiple correlation between several principal components and the criterion and R_{g+s}^c is the corrected for range restriction multiple correlation.

The stepwise regressions in the corrected matrices, the superior parameter estimates, revealed a situation much closer to previous findings. Psychometric **g** entered 24 times out of 32 regression analyses of four criteria on eight jobs. Twenty-three times **g** entered first. The reason for the disparity between these findings and the findings in the uncorrected (incorrect) correlations are the artifactual nature of observed correlations (Hunter, Schmidt, & Jackson, 1982). Stepwise regression methods begin by determining the highest correlation among the predictors with the criterion. When the highest correlation is different in the selected sample than in the population, the order of regression will change. Just as artifacts cloud the interpretation of observed correlations, unless subjected to proper estimation correction techniques, so too the observed uncorrected correlations reported here are liable to misinterpretation.

<u>Hands-on-work samples</u>. Using **HOPT** as the criterion and computing stepwise regressions using the uncorrected correlations, five of the eight jobs showed significant prediction by ability. In these regressions, **g** was the predictor to enter first three times and specific abilities added only to two of these three. Psychometric **g** entered the regression equation in 2nd order for one job.

Without corrections for artifactual depression, the average correlation of **g** and the **HOPT** criterion was .30 for the four jobs. In the two instances where **s** added to **g** the average increment was .12. Across all four jobs, the average increment to **g** was .06.

When **HOPT** was regressed against the predictors using the correlation matrix corrected for restriction of range, six of eight jobs showed correlations of ability and **HOPT**. In five of the six regressions, **g** entered first with an average simple correlation of .51. In four AFSCs with the data fully corrected, **s** added an average .08 to the prediction. For all six jobs where aptitude predicted job performance e.g., including the two where **s** added nothing, the average addition dropped to .05.

One job, Information Systems Radio Operator, showed a significant prediction of **HOPT** by the third principal component ($r = -.30$) in the uncorrected data. In the corrected data principal component three enters first ($r = -.29$) and **g** increases the prediction ($R = .40$). Psychometric **g** was the best predictor for **HOPT**.

<u>Interview Testing</u>. When **INT** was the criterion in the uncorrected regressions, **g** entered the stepwise procedure first three times in six jobs when aptitude predicted job performance. The average correlation of **g** in these three was .30. Specific or invested abilities added nothing in these three cases.

In three jobs, specific or invested abilities entered the regressions first. These were not the same **s**s in each case. On average the correlation of specific ability and job performance was .32.

In corrected regressions, there was significant prediction of the criterion and **g** entered firs for each job with an average correlation of .45. Measures of specific or invested abilities added six times (out of seven) with an average increment of .08.

Principal component two was the only significant predictor of **INT** for the job of Air Crew Life Support Specialist in the uncorrected and corrected analyses. The correlations were -.28 and -.26, respectively. Psychometric **g** was the best predictor of the technical interviews.

Walk Through Performance Test. The **WTPT** is a combination of the **HOPT** and the **INT** and provides better content sampling of the tasks in the jobs. When stepwise regressions were computed in the uncorrected data, **g** entered in six jobs (five times in first order; average r = .27) when there was prediction of the criterion. The average correlation of **g** and the criteria was .27 in all jobs where significant prediction occurred and **s** added to **g** in four jobs for an average increase in predictive efficiency of .13 on those four jobs and .02 when all six significantly predicted jobs were included in the average.

Computing the regressions in the fully corrected data there was significant prediction for seven of the eight jobs and **g** entered first in all seven for an average correlation of .47. Specific or invested abilities improved prediction in five jobs with an average increase of .09. The average gain for the seven jobs was .06. As before, psychometric **g** was the best predictor of **WTPT**.

The job performance **WTPT** criterion for Aircrew Life Support Specialist was not predictable by aptitude.

Supervisory Task Ratings. Analyses of **SUPR** were conducted in the same way as the other criteria. There was significant prediction of the criteria in the uncorrected data for only one job, Information Systems Radio Operator. The correlation, .28, was due to **g**. The measures of s did not add to the regression.

In the corrected analyses, five jobs were significantly predicted and in four of these **g** entered and in each case, first. In the fifth job, Jet Engine Mechanic, only the fourth principal component was a significant predictor. The average correlation of **g** for the four jobs was .34 and for one job, Avionics Communications Specialist, **s** added an increment of .19 to the prediction afforded by **g**.

In three jobs, the criterion of **SUPR** could not be predicted. These were Precision Measuring Equipment Specialist, Aerospace Ground Equipment Specialist, and Personnel Administration Specialist.

The corrected analyses provided the best estimates of the correlations in the population. They therefore present the best expressions of the relationships between the criteria and the predictors. There were 32 regressions calculated (8 jobs x 4 criteria = 32) and in 26 of these, significant prediction occurred. In these 26, **g** entered stepwise regressions first 23 (88%) times (See Table 5). In two of the remaining three cases, **g** was not predictive.

Sixteen times in the 23 regressions where g entered first other principal components added to the regression. The average corrected-for-cross-validation (Kennedy, 1988) increase in R due to s adding to g when g entered first was .06.

Table 5. Count of Principal Components Entering Regression Equations by Criterion Measure

Frequency of Entry of Principal Components

	Component									
Criterion	1	2	3	4	5	6	7	8	9	10
HOPT	6	2	1	1	2	0	0	1	0	0
INT	7	4	2	0	2	0	0	1	0	0
WTPT	7	2	2	1	1	0	0	1	0	0
SUPR	5	0	0	2	0	1	0	1	0	0

Note. The first principal component is g. Frequencies are based on regressions using correlation matrixes corrected for restriction of range.

In all 24 instances where g significantly entered the regression, regardless of order, the average correlational increase using g plus s was .06 corrected for cross-validation by Stein's operator (Kennedy, 1988).

Except for g, interpretation of the principal components which entered the regressions was difficult. In general, there was little similarity of which principal components were predictive for which jobs. For example, the two jobs which the Air Force uses G to select had different components potent in prediction. For the two E jobs, the only common principal component was g. Principal component 4 added to the prediction afforded by g for one job and principal components 3, 6, and 8 for the other. Much the same was found for the A jobs with g the common predictor and principal component 3 adding to prediction in one job and principal component 2 in another.

The two M jobs had similar patterns of prediction by the principal components. In both cases, principal components 2 and 5 increased prediction beyond g. Principal component 2 is reasonably a surrogate for gender (Jones, 1988) with its negative weighting on technical subtests and its positive weighting on speeded subtests. It separates those who perform well in subtests such as Auto & Shop Information (males) and Electronics Information (males) from those who perform well in the two speeded subtests (females). Principal component 5 was less interpretable but separated those who did well in one speeded subtest, simple arithmetic--Numerical Operation, from those who did well in the other speeded subtest, selecting from a table the number associated with a word--Coding Speed.

In previous studies (Ree & Earles, 1990; 1991b) when the principal components were used to predict training grades, it was found that g was the most potent predictor and that the specific or invested abilities added little to prediction. The same was true for predicting the job performance criteria but not as strongly as for training criteria.

For these job performance criteria, the potential exists that specific or invested abilities will offer classification utility. Studies to illuminate the theoretical and practical consequences of optimum classification such as sex bias, ethnic bias, adverse impact, regression effects, with-in job ability distributions, and individual vocational interest need to be accomplished.

References

Brogden, H. E. (1946). On the interpretation of the correlation coefficient as a measure of predictive efficiency. Journal of Educational Psychology, 37, 65-76.

Brogen, H. E. (1951). Increased efficiency of selection resulting from replacement of a single predictor with several different predictors. Journal of Educational Psychology, 11, 173-196.

Cattell, R. (1971). Abilities: Their structure, growth, and action. Boston: Houghton Mifflin.

Cattell, R. (1987). Intelligence: Its structure, growth, and action. Netherlands: Elsevier Science Publishing.

Christal, R. E. (1974). The United States Air Force occupational research project (AFHRL-TR-73-75). Brooks AFB TX: Occupational Research Division, AIr Force Human Resources Laboratory.

Department of Defense (1984). Test Manual for the Armed Services Vocational Aptitude Battery. North Chicago, IL: United States Military Entrance Processing Command.

Earles J. A., & Ree, M. J. (in press). The predictive validity of the ASVAB for training grades. Educational and Psychological Measurement.

Gottfredson, L. S. (1986). Foreword, The g factor in employment. Journal of Vocational Behavior, 29, 293-296.

Hedge J. W., Dickinson, T. L., & Bierstedt, S. A. (1988). The use of videotape technology to train administrators of Walk-Through Performance Testing (AFHRL-TP-87-71). Brooks AFB TX: Training Systems Division, Air Force Human Resources Laboratory.

Hedge J. W., & Lipscomb M. S. (1987). Walk-Trough Performance Testing (AFHRL-TP-87-8). Brooks AFB TX: Training Systems DIvision, Air Force Human Resources Laboratory.

Hedge J. W., Lipscomb M. S., & Teachout, M. S. (1988). Work sample testing in the Air Force job performance measurement project. In M. S. Lipscomb & J. W. Hedge (Eds.), Job performance measurement: Topics in the performance measurement of Air Force enlisted personnel (AFHRL-TP-87-58). Brooks AFB TX: Air Force Human Resources Laboratory.

Hedge, J. W., & Teachout, M. S. (1986). Job performance measurement: A systematic program of research and development. In R. J. Vance (Chair), First results of the U. S. Air Force performance project. A Symposium conducted at the annual meeting of the American Psychological Association, Atlanta, GA.

Hotelling, H. H. (1933a). Analysis of a complex of statistical variables with principal components. Journal of Educational Psychology, 24, 417-441.

Hotelling, H. H. (1933b). Analysis of a complex of statistical variables with principal components (continued). Journal of Educational Psychology, 24, 498-520.

Hull, C. (1928). Aptitude testing. Great Britain: World Book.

Hunter, J. E.(1983). Validity generalization of theASVAB: Higher validity for factor analytic composites. Rockville, MD: Research Applications, Inc.

Hunter, J. E. (1984a). The prediction of job performance in the civilian sector using the ASVAB. Rockville, MD: Research Applications, Inc.

Hunter, J. E. (1984b). The validity of the ASVAB as a predictor of civilian job performance. Rockville, MD: Research Applications, Inc.

Hunter, J. E. (1984c). The validity of the Armed Services Vocational Aptitude Battery (ASVAB) high school composites. Rockville, MD: Research Applications, Inc.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. Journal of Vocational Behavior, 29, 340-362.

Hunter, J. E., Crosson, J. J., & Friedman, D. H. (1985). The validity of the Armed Services Vocational Aptitude Battery (ASVAB) for civilian and military job performance. Rockville, MD: Research Application, Inc.

Hunter, J. E., Schmidt, F., & Jackson, G. (1982). Meta-analysis: Cumulating research findings across studies. Beverly Hills: Sage.

Jensen, A. R. (1980). Bias in mental testing. New York: The Free Press.

Jensen, A. R. (1987). Editorial: Psychometric g as a focus of concerted research effort. Intelligence, 11, 193-198.

Jones, G. E. (1988). Investigation of the efficacy of general ability versus specific ability as predictors of occupational success. Unpublished master's thesis, St. Mary's University, San Antonio, TX.

Kendall, M., Stuart, A., & Ord, J. K. (1983). The advanced theory of statistics. Volume 3, (4th ed), New York: Macmillan.

Kennedy, E. (1988). Estimation of the squared cross-validity coefficient in the context of best subset regression. Applied Psychological Measurement, 12, 231-237.

Lipscomb, M. S. (1984). A task-level domain sampling strategy: A content valid approach. In S. Zedeck (Chair), Walk-Through Performance Testing: An innovative approach to work sample testing. A Symposium at the annual meeting of the American Psychological Association, Toronto Canada.

Lipscomb, M. S., & Dickinson, T. L. (1988). The Air Force domain specification and sampling plan. In M. S. Lipscomb & J. W. Hedge (Eds.), Job performance: Topics in the performance measurement of Air Force enlisted personnel (AFHRL-TP-87-58). Brooks AFB TX: Training Systems Division, Air Force Human Resources Laboratory.

Maier, M. H. & Sims, W. H. (1986). The ASVAB score scales: 1980 and World War II CNR 116, Alexandria, VA: Center for Naval Analyses.

McIntyre, R. M., Smith D. E., & Hassett C. E. (1984). Accuracy of performance ratings as effected by rater training and perceived purpose of rating. Journal of Applied Psychology, 69, 147-156.

Palmer, P., Hartke, D. D., Ree, M. J., Welsh, J. R., & Valentine, L. D., Jr. (1988). Armed Services Vocational Aptitude Battery (ASVAB): Alternate forms reliability (Forms 8, 9, 10, and 11) (AFHRL-TP-87-48), Brooks AFB,TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Raven, J. (1938). Progressive matrices: A perceptual test of intelligence. London: H. K. Lewis.

Ree, M. J. & Earles, J. A. (1990). The differential validity of a differential aptitude test (AFHRL-TR-89-59). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Ree, M. J., & Earles, J. A. (1991a). The stability of g across different methods of estimation. Intelligence, 15, 271- 278.

Ree, M. J., & Earles, J. A. (1991b). Predicting training success: Not much more than g. Personnel Psychology, 44, 321-332.

Ree, M., & Wegner, T. G. (1990). Correcting differences in answer sheets for the 1980 Armed Services Vocational Aptitude Battery reference population. Military Psychology, 2, 157-169.

Schmid, J., & Leiman, J. M. (1955). The development of hierarchical factor solutions. Psychometrika, 22, 53-61.

Spearman, C. (1904). "General Intelligence," objectively determined and measured. American Journal of Psychology, 15, 201-293.

Stewart, N. (1947). Scores of Army personnel grouped by occupation. Occupations, 26, 1-37.

Thurstone, L. L. (1938). Primary mental abilities. Chicago: University of Chicago Press.

Vernon, P. (1960). Intelligence and attainment tests. New York: Philosophical Library.

Ward, J. H., & Jennings, E. (1973). Introduction to linear models. Englewood Cliffs, New Jersey: Prentice-Hall.

Welsh, J. R., Jr., Kucinkas, S., & Curran, L. (1990). Armed Services Vocational Aptitude Battery (ASVAB): Integrative review of validity studies (AFHRL-TR-90-22). Brooks AFB TX: Air Force Human Resources Laboratory.

Welsh, J. R., Jr., Trent, T., Nakasone, R., Fairbank, B., Kucinkas, S., & Sawin, L. (1990). Annotated bibliography of Armed Services Vocational Aptitude Battery (ASVAB) Validity Studies (AFHRL-TR-89-76). Brooks AFB TX: Air Force Human Resources Laboratory.

Wigdor, A., & Green, B. F., Jr. (1987). Assessing the performance of enlisted personnel: Evaluation of a joint-service project. Washington DC: National Academy Press.

Wilbourn, J. M., Valentine, L. D., Jr., & Ree, M. J. (1984). Relationships of the Armed Services Vocational Aptitude Battery (ASVAB) forms 8, 9, and 10 to Air Force technical school final grades (AFHRL-TR-84-8). Brooks AFB TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.